**Data Sources:**

* **yfinance:** Provided historical stock price and volume data.
* **Alpha Vantage:** Provided sentiment data.

**Models Used:** Logistic Regression, Gradient Boosting, Random Forest.

**Challenges Faced:**

1. **Limited Sentiment Data:** The sentiment data from Alpha Vantage was sparse insufficient for capturing the real-time sentiment shifts that can significantly influence stock prices. This limitation was a key factor in the underperformance of the models. Real-time data from social media platforms like X (which was not used due to pricing) could have offered more insights into the investor behavior.
2. **Forward Filling for Sentiment:** To address gaps in the sentiment data, forward filling was applied to maintain data consistency. While this technique ensured no missing values, it may have led to overestimating the impact of old sentiment data, causing the models to misinterpret market sentiment changes.

**Model Performance:** The performance of all models was relatively poor across most stocks, with only a few isolated cases (TNXP using Logistic Regression) achieving moderate accuracy.

* **APLD Stock:** Logistic Regression and Gradient Boosting both performed equally (50% accuracy).
* **CNEY Stock:** Logistic Regression slightly outperformed Gradient Boosting (59.4% vs. 46.9% accuracy).
* **KTTA Stock:** Both models underperformed, with Gradient Boosting achieving 47.1% accuracy, while Logistic Regression scored 41.2%.
* **ONCO Stock:** Logistic Regression marginally outperformed Gradient Boosting (58.3% vs. 54.2% accuracy).
* **TNXP Stock:** Logistic Regression performed significantly better (80% accuracy), while Gradient Boosting struggled (40%).
* **Random Forest Performance:**
  + **Accuracy:** 53.7%
  + **Precision:** 0.48 (class 0), 0.58 (class 1)
  + **Recall:** 0.48 (class 0), 0.58 (class 1)
  + **F1-Score:** 0.48 (class 0), 0.58 (class 1)